

AI Choreographer: Music Conditioned 3D Dance Generation with AIST++

Year: 2021 | Citations: 613 | Authors: Ruilong Li, Sha Yang, David A. Ross, Angjoo Kanazawa

Abstract

We present AIST++, a new multi-modal dataset of 3D dance motion and music, along with FACT, a Full-Attention Cross-modal Transformer network for generating 3D dance motion conditioned on music. The proposed AIST++ dataset contains 5.2 hours of 3D dance motion in 1408 sequences, covering 10 dance genres with multi-view videos with known camera poses—the largest dataset of this kind to our knowledge. We show that naively applying sequence models such as transformers to this dataset for the task of music conditioned 3D motion generation does not produce satisfactory 3D motion that is well correlated with the input music. We overcome these shortcomings by introducing key changes in its architecture design and supervision: FACT model involves a deep cross-modal transformer block with full-attention that is trained to predict N future motions. We empirically show that these changes are key factors in generating long sequences of realistic dance motion that are well-attuned to the input music. We conduct extensive experiments on AIST++ with user studies, where our method outperforms recent state-of-the-art methods both qualitatively and quantitatively. The code and the dataset can be found at: <https://google.github.io/aichoreographer>.